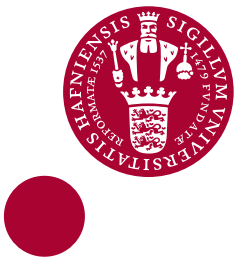

Systems genetics analysis of obesity using RNA-Seq data in an F2 pig resource population

Kogelman, Lisette; Zhernakova, D. V.; Westra, H.-J.; Cirera Salicio, Susanna; Fredholm, Merete; Franke, L.; Kadarmideen, Haja

Published in:
Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Publication date:
2014

Document version
Early version, also known as pre-print

Systems genetics analysis of obesity using RNA-Seq data in an F2 pig resource population

L.J.A. Kogelman[1], D.V. Zhernakova[2], H-J. Westra[2], S. Cirera[1], M. Fredholm[1], L. Franke[2] and H. N. Kadarmideen[1]*
*(*Correspondence)*
[1]Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark [2]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands

ABSTRACT: Obesity is a complex problem, associated with many diseases. Microarray gene expression data have been extensively used to detect differentially expressed (DE) genes and expression quantitative trait loci (eQTL), however, RNA-Sequencing data have the potential to reveal novel genes involved in complex traits. The objective was to elucidate biological pathways and potential biomarkers for obesity in a porcine model, by systems genetics approaches using RNA-Sequencing data. Previously, we created an F2 pig population which was deeply phenotyped and genotyped. Based on their degree of obesity, 36 animals were selected for RNA-Sequencing. Analysis included DE, pathway detection and eQTL mapping. We identified 198 DE genes, which could be divided in immune and developmental related processes. Furthermore, we revealed 761 cis-eQTLs of which several could be linked to obesity. Concluding, systems genetics analysis of RNA-Seq data elucidated biologically relevant pathways and potential genetic biomarkers affecting obesity.

Introduction

Obesity is a complex health problem, associated with several metabolic diseases. The exponential rise in the incidence of obesity worldwide, and its huge welfare, social and economic impact has enlarged the urge of gaining knowledge on the biologic and genetic background. Here, we use a porcine model for human obesity. It has been shown that the pig has similar metabolic, digestive and cardiovascular features, and it resembles humans more than rodents (Spurlock and Gabler (2008)). We previously created an F2 pig population, and revealed the potential of this population to study human obesity (Kogelman et al. (2013)).

Currently, Next Generation Sequencing (NGS) technologies are offering a huge potential for studying complex traits and diseases. RNA-Sequencing (RNA-Seq) data is replacing microarray expression studies, because of their huge potential in e.g., more precise measurements of expression levels and the potential to discover novel transcripts (Wang et al. (2009)). Sequencing-based technologies have been shown to give promising results in e.g., kidney disease (Mimura et al. (2013)), and a huge potential is expected for other complex, multifactorial diseases and traits, such as obesity.

To elucidate the biological and genetic background of complex diseases, several network approaches have been used to detect pathways and potential causal genes. Furthermore, previous studies have shown that studying expression quantitative trait loci (eQTLs) is an appropriate way of increasing the knowledge of complex diseases and traits (Morley et al. (2004), Kadarmideen et al. (2006)). In eQTL mapping each transcript abundance is treated as a phenotype and typical QTL mapping or GWAS approaches are applied to this expression phenotype or expression-trait (see Kadarmideen et al. (2006) for definitions).

The objective of this study was to increase the systematic understanding of the transcriptional (co-) regulations of obesity and related traits by detecting differentially expressed genes and eQTLs. This will result in the interpretation of the functions of known and novel genes associated with obesity, and increase the understanding of obesity-related biological pathways.

Materials and Methods

Experimental design. An F2 pig population was created using Danish production pig breeds (i.e. Yorkshire and Duroc sows) and Göttingen minipig boars. The production breeds are intensively selected for leanness and growth, while the Göttingen minipigs are prone to obesity and share metabolic impairments seen in obese humans (Johansen et al. (2001)). As published earlier, the F2 population (454 pigs) was intensively phenotyped for, e.g. weight, body confirmation, DXA scanning, and slaughter characteristics (Kogelman et al. (2013)). All animals were genotyped using the Illumina Porcine 60K SNP Chip.

Obesity Index. In animal breeding, multi-trait selection indexes are used to select animals based on estimated breeding values (EBVs) for several traits of interest (Cameron (1997)). Based on this, we created the Obesity Index (OI), by calculating selection index weights and combining EBVs for nine different obesity-related traits (reported in Kogelman et al. (2013)) into one aggregate total merit index for all animals. Traits selected for the obesity index were: weight and abdominal circumference at slaughter age, average daily gain, estimated fat mass and percentage of fat at DXA scanning, back fat thickness at position 1 and position 2, weight of leaf fat and weight of omental fat at slaughtering. This resulted in a single genetic OI score, with the potential to determine whether an animal was genetically obese or lean.

RNA Sequencing. Based on the OI 36 (12 high, 12 intermediate and 12 low OI) animals were selected for RNA-Seq. RNA-Seq was performed on the Illumina HiSeq2500 (AROS, Denmark), using subcutaneous fat tissue. Alignment was performed using STAR, resulting in (after quality control) approximately 30 million mapped reads per sample, mapped to 25,322 unique transcripts. After alignment, raw counts were normalized and corrected for gender.

Differential expression. Differentially expressed (DE) transcripts for the OI were detected using a linear model, fitting the OI as continuous variable and gender as covariate using the R-package Limma (Smyth (2005)). Estimates were corrected for multiple testing using the Benjamini & Hochberg (FDR) correction. Transcripts were detected as significantly DE using an FDR < 0.05.

Functional annotation. The associated genes at the DE transcripts were detected using BioMart (Durinck et al. (2005)). To identify overrepresented Gene Ontology (GO) terms and pathways we used the online available database GeneNetwork (http://www.genenetwork.nl), which is constructed using human, mouse and rate expression data. Gene functions were predicted against known pathways and gene sets in various biological databases. Overrepresentation of GO-terms and pathways was tested using the Mann-Whitney U test, and P-values were afterwards corrected for multiple testing using the Bonferroni correction.

eQTL studies. eQTL mapping was performed to find the downstream effects of the genetic variants associated with obesity, by treating each transcript abundance as phenotype. RNA-Seq data was corrected for the gender effect, normalization was applied, and technical variation was checked using a principal component (PC) analysis on the RNA-Seq sample correlation matrix (Westra et al. (2013)). Cis-eQTLs were defined when the associated SNP was within 10 Mb of the transcript. This window was chosen based on preliminary results, the large haplotype block size of pigs (on average 400 Kb), and the knowledge that haplotype block sizes in F2 populations will be even larger. Only SNPs with a minor allele frequency of >0.05 and a Hardy-Weinberg equilibrium P-value of >0.001 were included in the analyses. Associations were determined using Spearman's rank correlation and to correct for multiple testing, the analysis was repeated 10 times, each time permuting the sample labels, in order to retain the correlation structure within the genotype and gene-expression data. The resultant p-value distribution was applied as a null-distribution to control the false discovery rate at 0.05.

Results and Discussion

Differential expression. We detected a total of 189 DE genes (FDR<0.05), of which 139 were corresponding to unique genes.

Table 1. Top 10 differentially expressed transcripts

Transcript	Expr.	P-value	Assoc Gene
ENSSSCG00000025188	22.01	$1.8E^{-3}$	LEPR
ENSSSCG00000028062	20.70	$1.8E^{-3}$	TCEAL3
ENSSSCG00000012566	23.63	$3.3E^{-3}$	
ENSSSCG00000003341	25.65	$4.3E^{-3}$	TAS1R3
ENSSSCG00000007005	27.52	$4.6E^{-3}$	CSGALNACT1
ENSSSCG00000016093	24.62	$7.3E^{-3}$	
ENSSSCG00000004177	32.19	$7.6E^{-3}$	RPS12
ENSSSCG00000022685	24.57	$7.6E^{-3}$	ROM1
ENSSSCG00000009060	24.96	$1.9E^{-2}$	MAML3
ENSSSCG00000016957	21.96	$1.9E^{-2}$	CD180

The most highly DE gene was the LEPR (Leptin Receptor) gene. Leptin is a hormone produced by adipose tissue and plays a key role in energy regulation and appetite. Dysfunctioning of the LEPR gene results in obesity in human and mouse models (Gilbert et al. (2003)). The TAS1R3 gene encodes a taste receptor. The perception of taste has been linked before to eating behavior (Grimm and Steinle (2011)) and therefore also associated with obesity. The RPS12 gene is encoding the Ribosomal Protein S12, which has been shown to be associated with diabetic nephropathy in African Americans (McDonough et al. (2001)). The MAML3 gene is encoding the mastermind-like 3 protein, which has been associated with protein intake in a meta-analysis. However, it did not pass the significance threshold in the 2nd stage (Tanaka et al. (2013)). The last gene in the top 10 DE genes is the CD180 gene, previously called Ly64. This gene has been shown to be differentially expressed in the liver of mice, for a subphenotype of diabetes (Mir et al. (2003)). Moreover, it was also differentially expressed in mice showing diabetic development (Fornari et al. (2011)). Furthermore, we detected the RPL13A gene (P-value=0.02), of which disruption caused resistance to lipotoxity (Michel et al. (2011)) and the TNMD gene (P-value=0.02), previously associated with adiposity, glucose metabolism and in men also type 2 diabetes development (Tolppanen et al. (2007)).

Functional annotation. A network was constructed using the GeneNetwork software on all DE genes.

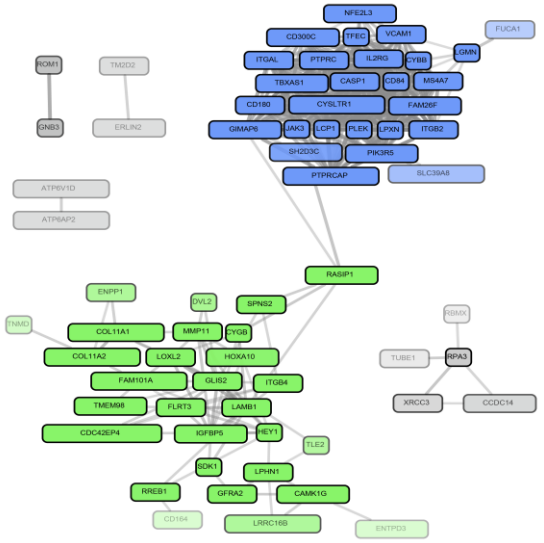

Figure 1. Network construction of DE genes, with the clustering of two subnetworks (blue and green). The intensity of the colors represents the significance of presence of the gene in the particular network structure.

The constructed network of DE genes shows two clear clusters of highly interconnected genes. Those two clusters, or subnetworks, were separately analyzed using GeneNetwork.

The first subnetwork (blue in Fig. 1) shows a strong association with immune related pathways. Highly significant GO terms in the Biological Processes are, for example, *leukocyte migration* ($P_{adj}=6.62E^{-14}$), *interleukin-12*

production (P_{adj}=7.45E^{-14}) and *cytokine production* (P_{adj}=7.50E^{-14}). The most significant KEGG pathway is the *chemokine signaling pathway* (P_{adj}=3.46E^{-15}). This pathway has been associated with the inflammation response in obesity, and subsequent development of insulin resistance (Tsuguhito (2013)). Moreover, in the Mouse Genome Informatics (MGI) database an *increased circulating tumor necrosis factor level* (P_{adj}=2.18E^{-14}) was highly overrepresented, which has repeatedly shown to be associated with obesity-linked insulin resistance (Hotamisligil et al. (1993)).

The second subnetwork (green in Fig. 1) showed a strong association with developmental associated pathways. The most highly overrepresented GO term in the Biological Processes was *regulation of canonical Wnt receptor signaling pathway* (P_{adj}=1.05E^{-11}), and the *Wnt signaling pathway* is also the most overrepresented KEGG pathway (P_{adj}=2.78E^{-6}). This pathway has a major role in the development of obesity, by generation of new adipocytes (adipogenesis) (Laudes (2011)).

eQTL study. Genome-wide genetic analyses of transcriptomic variation measured by RNA-Seq resulted in the detection of eQTLs. In total 761 cis-eQTLs were detected, considering all SNPs and transcripts (Fig. 2A).

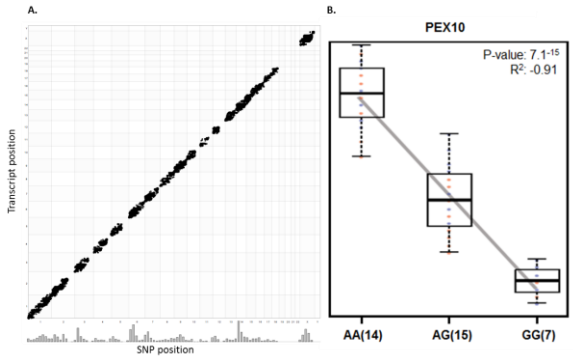

Figure 2. eQTL study results. The dotplot (A) shows the detected eQTLs with their SNP and transcript position. The boxplot (B) shows one of the highly significant eQTLs: the PEX10 gene.

The PEX10 gene was one of the most significant eQTLs detected (Fig. 2B). This gene is involved in the import of peroxisomal matrix proteins, which has an important role in the breakdown of very long chain fatty acids (Chen et al. (2010).

Future studies will include further investigation of the DE genes, for example, by investigating the up- and down-regulated genes. Moreover, as the eQTL approach does not take the family structure of the pigs into account, the eQTL findings will be validated using a linkage based method used in R/QTL. Finally, both DE and eQTL findings will be validated in human expression data.

Conclusion

We used systems genetics approaches, i.e. DE, functional annotation and eQTL studies, to elucidate the systems biological and genetic background of obesity and obesity-related diseases. In total 189 DE genes were detected and many could be linked to human obesity, emphasizing the validity of the pig as a model for human obesity. A network analyses on DE genes revealed two subnetworks of closely related genes; those subnetworks represent different kind of biological processes directly or indirectly related to obesity, e.g. immune related processes and developmental processes. Furthermore, the detection of various eQTLs gives us the opportunity to reveal whole-genome regulatory mechanisms and potential causal genes for human obesity and obesity-related diseases.

Acknowledgements

The project is supported by a grant from the Ministry of Science and Technology to the "UNIK Project for Food Fitness and Pharma for Health", funding from the Danish Council for Strategic Research to BioChild Project, and from a Ph.D. stipend awarded to Lisette J.A. Kogelman from University of Copenhagen. Authors thank EU-FP7 Marie Curie Actions – Career Integration Grant (CIG-293511) granted to Haja N. Kadarmideen for funding this study.

Literature Cited

Cameron N.D. (1997). CABI, ISBN-10: 0851991696.
Chen H., Liu Z. and Huang X. (2012). Hum. Mol. Genet. 19(3): 494-505
Durinck S., Moreau Y., Kasprzyk A. et al. (2005). Bioinf. 21(16): 3439-3440.
Fornari, T.A., Donate, C, Macedo, E.T, et al. (2011). Clin. and Dev. Imm., 2011: 12.
Gilbert, M., Magnan, S., Turban S., et al. (2003). Diabetes, 52(2): 277-282
Grimm E.R. and Steinle N.I. (2011). Nutr Rev. 69(1): 52-60.
Hotamisligil G.S., Shargill N.S. and Spiegelman B.M. (1993). Science 259(5091):87-91
Johansen, T., Hansen H.S., Richelsen, B., et al (2001). Comp. Medicine, 51(2): 150-155.
Kadarmideen H.N., von Rohr P., Janss L.L.G. (2006). Mamm. Gen. 17(6):548-564
Kogelman L.J.A., Kadarmideen H.N., Mark T., et al (2013). Frontiers in Genetics, 4(29).
Laudes M. (2011). J. Mol. Endocrinol. 46:65-72.
McDonough, C.W., Palmer, N.D., Hicks, P.J., et al (2011). Kidney Int, 79(5): 563-572
Michel C.I., Holley C.L., Scruggs B.S., et al. (2012). Cell Metab. July 6;14(1):33-44
Mimura I., Kanki Y., Kodama T., et al. (2013). Kidney Int. 85, 31-38.
Mir, A.A., Myakishev, M.V., Polesskaya, O.O., et al. (2003). Genomics, 81(4): 378-390
Morley M., Molony C.M., Weber T.M., et al. (2004). Nature, 430: 743-747
Smyth G.K. (2005) Springer New York p.397-420.
Spurlock M.E. and Gabler N.K. (2008). J. Nutr. 138:397-402.
Tolppanen A.M., Pulkkinen L., Kolehmainen M., et al. (2007). Obesity, 15(5): 1082-8
Tsuguhito O. (2013). Diabetes Metab J. 37(3): 165-172.
Wang Z., Gerstein M., and Snyder M. (2009). Nat Rev Genet. 10(1):57-63.
Westra H-J., Peters M.J., Esko T., et al. (2013). Nat. Gen. 45-1238-1243